

FAMILY 9 ENDO-BETA-1,4-GLUCANASES

The present invention relates to an enzyme exhibiting endo- β -1,4-glucanase activity which enzyme belongs to family 9 of glycosyl hydrolases and is at least 75% homologous to a
5 *Bacillus licheniformis* family 9 endo- β -1,4-glucanase, to an isolated polynucleotide molecule encoding such an endo- β -1,4-glucanase, and use of the enzyme in the detergent, paper and pulp, oil drilling, oil extraction, wine and juice, food ingredients, animal feed or textile industries.

10

BACKGROUND OF THE INVENTION

Cellulose is a polymer of glucose linked by β -1,4-glucosidic bonds. Cellulose chains form numerous intra- and intermolecular hydrogen bonds, which result in the formation of
15 insoluble cellulose microfibrils. Microbial hydrolysis of cellulose to glucose involves the following three major classes of cellulases: (i) endoglucanases (EC 3.2.1.4) which cleave β -1,4-glucosidic links randomly throughout cellulose molecules; (ii) cellobiohydrolases (EC 3.2.1.91) which digest cellulose
20 from the nonreducing end, releasing cellobiose; and (iii) β -glucosidases (EC 3.2.1.21) which hydrolyse cellobiose and low-molecular-mass cellodextrins to release glucose.

Cellulases are produced by many microorganisms and are often present in multiple forms. Recognition of the economic
25 significance of the enzymatic degradation of cellulose has promoted an extensive search for microbial cellulases which can be used industrially. As a result, the enzymatic properties and the primary structures of a large number of cellulase have been investigated. On the basis of the results of a hydrophobic
30 cluster analysis of the amino acid sequence of the catalytic domain, these cellulases have been placed into different families of glycosyl hydrolases; fungal and bacterial glycosyl hydrolases have been grouped into 35 families (Henrissat et. al. (1991), (1993)). Most cellulases consist of a cellulose-binding
35 domain (CBD) and a catalytic domain (CAD) separated by a linker which may be rich in proline and hydroxy amino residues. Another

classification of cellulases has been established on the basis of the similarity of their CBDs (Gilkes et al. (1991)) giving five families of glycosyl hydrolases (I-V).

Cellulases are synthesized by a large number of microorganisms which include fungi, actinomycetes, myxobacteria and true bacteria but also by plants. Especially endo- β -1,4-glucanases of a wide variety of specificities have been identified. Many bacterial endoglucanases have been described (Henrissat (1993); Gilbert et al., (1993)).

10 An important industrial use of cellulolytic enzymes is the use for treatment of paper pulp, e.g. for improving the drainage or for deinking of recycled paper. Another important industrial use of cellulolytic enzymes is the use for treatment of cellulosic textile or fabric, e.g. as ingredients in detergent compositions or fabric softener compositions, for bio-polishing of
15 new fabric (garment finishing), and for obtaining a "stone-washed" look of cellulose-containing fabric, especially denim, and several methods for such treatment have been suggested, e.g. in GB-A-1 368 599, EP-A-0 307 564 and EP-A-0 435 876, WO
20 91/17243, WO 91/10732, WO 91/17244, PCT/DK95/000108 and PCT/DK95/00132.

There is an ever existing need for providing novel cellulase enzymes or enzyme preparations which may be used for applications where cellulase, preferably an endo- β -1,4-glucanase, activity (EC 3.2.1.4) is desirable.
25

The object of the present invention is to provide novel enzymes and enzyme compositions having substantial cellulolytic activity under slightly acidic to alkaline conditions and improved performance in paper pulp processing, textile
30 treatment, laundry processes, extraction processes or in animal feed; preferably such novel well-performing endoglucanases are producible or produced by using recombinant techniques in high yields.

35 SUMMARY OF THE INVENTION

The inventors have found a novel enzyme having substantial cellulolytic activity, i.e. an endo- β -1,4-glucanase (classified

according to the Enzyme Nomenclature as EC 3.2.1.4), which enzyme is endogenous to *Bacillus licheniformis* and belongs to family 9 of glycosyl hydrolases, and the inventors have succeeded in cloning and expressing a DNA sequence encoding such an enzyme.

Accordingly, in its first aspect the present invention relates to an enzyme exhibiting endo- β -1,4-glucanase activity (EC 3.2.1.4) which is selected from one of (a) a polypeptide encoded by the DNA sequence of positions 76 to 1455 of SEQ ID NO:1; (b) a polypeptide produced by culturing a cell comprising the sequence of SEQ ID NO:1 under conditions wherein the DNA sequence is expressed; (c) an endo- β -1,4-glucanase enzyme having a sequence of at least 75% identity to positions 26-485 of SEQ ID NO:2 polypeptide comprising an amino acid sequence derived from the amino acid sequence of positions 26-485 of SEQ ID NO:2 when identity is determined by GAP provided in the GCG program package using a GAP creation penalty of 3.0 and GAP extension penalty of 0.1; and (d) a polypeptide encoded by the endoglucanase encoding part of the DNA sequence obtainable from the plasmid in *Escherichia coli* DSM 12805. The enzyme of the invention is identified as belonging to family 9 of glycosyl hydrolases as defined by Henrissat et al.

In its second aspect the invention relates to an isolated polynucleotide molecule, preferably a DNA molecule, encoding the catalytically active domain of an enzyme exhibiting endo- β -1,4-glucanase activity which molecule is selected from the group consisting of (a) polynucleotide molecules comprising a nucleotide sequence as shown in SEQ ID NO:1 from nucleotide 76 to nucleotide 1455, (b) species homologs of (a); (c) polynucleotide molecules that encode a polypeptide that is at least 75% identical to the amino acid sequence of SEQ ID NO:2 from amino acid residue 26 to amino acid residue 485, and (c) degenerate nucleotide sequences of (a) or (b); preferably a polynucleotide molecule capable of hybridizing to a denatured double-stranded DNA probe under medium stringency conditions, wherein the probe is selected from the group consisting of DNA probes comprising the sequence shown in positions 76-1455 of SEQ ID NO:1 and DNA

probes comprising a subsequence of positions 76-1455 of SEQ ID NO:1 having a length of at least about 100 base pairs.

A plasmid pSJ1678 comprising a DNA sequence encoding the endoglucanase of the invention has been transformed into a
5 strain of the *Escherichia coli* which was deposited by the inventors according to the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig,
10 Federal Republic of Germany, on 14 May 1999 under the deposition number DSM 12805.

In its third, fourth and fifth aspect the invention provides an expression vector comprising a DNA segment which is e.g. a polynucleotide molecule of the invention; a cell comprising the DNA segment or the expression vector; and a method of
15 producing an enzyme exhibiting cellulolytic activity, which method comprises culturing the cell under conditions permitting the production of the enzyme, and recovering the enzyme from the culture.

20 In yet another aspect the invention provides an isolated enzyme exhibiting cellulolytic activity, characterized in (i) being free from homologous impurities and (ii) the enzyme is produced by the method described above.

Further, the present invention relates to the use of such
25 an enzyme or the enzyme preparation of the invention for industrial applications such as for the treatment of wooden pulp or degradation of biomass.

The invention also relates to an isolated substantially pure biological culture of the *Escherichia coli* strain
30 DSM 12805 harbouring the endoglucanase encoding DNA sequence cloned into plasmid pSJ1678 present in *Escherichia coli* DSM 12805 which is derived from a strain of the bacterial species *Bacillus licheniformis*, or any mutant of said *E.coli* strain.

The endoglucanase of the invention is advantageous in a
35 number of industrial applicaitons by having a high specific activity on CMC (endoglucanase) and, in contrast to most other endoglucanases, the enzyme of the invention is able to degrade

highly crystalline cellulose. Furthermore, this enzyme has its optimal temperature at 60°C and is fully active between pH 5.5 and 9.5. Accordingly, the enzyme of the invention can advantageously be used for total biomass degradation, which normally
5 would need both cellobiohydrolase(s) (which has very little activity on CMC) and endoglucanase(s).

DETAILED DESCRIPTION OF THE INVENTION

The term "glycosyl hydrolase family" as used herein has
10 been described in Henrissat, B. "A classification of glycosyl hydrolases based of amino-acid sequence similarities." Biochem. J. 280: 309-316 (1991); Henrissat, B., Bairoch, A. "New families in the classification of glycosyl hydrolases based on amino-acid sequence similarities. Biochem. J. 293: 781-788 (1993); Henris-
15 sat, B., Bairoch, A. "Updating the sequence-based classification of glycosyl hydrolases." Biochem. J. 316: 695-696 (1996); and- Davies, G., Henrissat, B. "Structures and mechanisms of glycosyl hydrolases." Structure 3: 853-859 (1995); all of which are incorporated by reference.

20 The term "functional enzymatic properties" as used herein is intended to mean physical and chemical properties of a polypeptide exhibiting one or more catalytic activities. Examples of functional enzymatic properties are enzymatic activity, specific enzymatic activity, relative enzymatic activity to the maximum
25 activity (measured as a function of either pH or temperature), stability (degradation of enzymatic activity over time), DSC melting temperature, N-terminal amino acid sequence, molecular weight (usually measured in SDS-PAGE), isoelectric point (pI).

In the present context the term "expression vector" denotes
30 a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that provide for its transcription. Such additional segments may include promoter and terminator sequences, and may optionally include one or more origins of replication, one or
35 more selectable markers, an enhancer, a polyadenylation signal, and the like. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both. The

expression vector of the invention may be any expression vector that is conveniently subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which the vector is to be introduced. Thus, the vector may be an
5 autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together
10 with the chromosome(s) into which it has been integrated.

The term "recombinant expressed" or "recombinantly expressed" used herein in connection with expression of a polypeptide or protein is defined according to the standard definition in the art. Recombinant expression of a protein is
15 generally performed by using an expression vector as described immediately above.

The term "isolated", when applied to a polynucleotide molecule, denotes that the polynucleotide has been removed from its natural genetic milieu and is thus free of other extraneous or
20 unwanted coding sequences, and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes
25 with which they are ordinarily associated, but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators. The identification of associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, Nature 316:774-78, 1985). The term "an iso-
30 lated polynucleotide" may alternatively be termed "a cloned polynucleotide".

When applied to a protein/polypeptide, the term "isolated" indicates that the protein is found in a condition other than its native environment. In a preferred form, the isolated pro-
35 tein is substantially free of other proteins, particularly other homologous proteins (i.e. "homologous impurities" (see below)). It is preferred to provide the protein in a greater than 40% pure form, more preferably greater than 60% pure form.

Even more preferably it is preferred to provide the protein in a highly purified form, i.e., greater than 80% pure, more preferably greater than 95% pure, and even more preferably greater than 99% pure, as determined by SDS-PAGE.

5 The term "isolated protein/polypeptide may alternatively be termed "purified protein/polypeptide".

The term "homologous impurities" means any impurity (e.g. another polypeptide than the polypeptide of the invention) which originate from the homologous cell where the polypeptide of the
10 invention is originally obtained from.

The term "obtained from" as used herein in connection with a specific microbial source, means that the polynucleotide and/or polypeptide produced by the specific source, or by a cell in which a gene from the source have been inserted.

15 The term "operably linked", when referring to DNA segments, denotes that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in the promoter and proceeds through the coding segment to the terminator

20 The term "polynucleotide" denotes a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and may be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic
25 molecules.

The term "complements of polynucleotide molecules" denotes polynucleotide molecules having a complementary base sequence and reverse orientation as compared to a reference sequence. For example, the sequence 5' ATGCACGGG 3' is complementary to 5' CCCGTGCAT 3'.
30

The term "degenerate nucleotide sequence" denotes a sequence of nucleotides that includes one or more degenerate codons (as compared to a reference polynucleotide molecule that encodes a polypeptide). Degenerate codons contain different
35 triplets of nucleotides, but encode the same amino acid residue (i.e., GAU and GAC triplets each encode Asp).

The term "promoter" denotes a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of transcription. Promoter sequences are commonly, but not always, found in the 5' non-coding regions of genes.

5 The term "secretory signal sequence" denotes a DNA sequence that encodes a polypeptide (a "secretory peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized. The larger peptide is commonly cleaved to remove the
10 secretory peptide during transit through the secretory pathway.

POLYNUCLEOTIDES:

Within preferred embodiments of the invention an isolated polynucleotide of the invention will hybridize to similar sized
15 regions of SEQ ID No. 1, or a sequence complementary thereto, under at least medium stringency conditions.

In particular polynucleotides of the invention will hybridize to a denatured double-stranded DNA probe comprising either the full sequence encoding for the catalytic domain of
20 the enzyme which sequence is shown in positions 76-1455 of SEQ ID NO:1 or any probe comprising a subsequence of SEQ ID NO:1 having a length of at least about 100 base pairs under at least medium stringency conditions, but preferably at high stringency conditions as described in detail below. Suitable experimental
25 conditions for determining hybridization at medium, or high stringency between a nucleotide probe and a homologous DNA or RNA sequence involves presoaking of the filter containing the DNA fragments or RNA to hybridize in 5 x SSC (Sodium chloride/Sodium citrate, Sambrook et al. 1989) for 10 min, and
30 prehybridization of the filter in a solution of 5 x SSC, 5 x Denhardt's solution (Sambrook et al. 1989), 0.5 % SDS and 100 µg/ml of denatured sonicated salmon sperm DNA (Sambrook et al. 1989), followed by hybridization in the same solution containing a concentration of 10ng/ml of a random-primed (Feinberg, A. P.
35 and Vogelstein, B. (1983) Anal. Biochem. 132:6-13), 32P-dCTP-labeled (specific activity higher than 1 x 10⁹ cpm/µg) probe for 12 hours at ca. 45°C. The filter is then washed twice for 30 minutes in 2 x SSC, 0.5 % SDS at least 60°C (medium stringency),

still more preferably at least 65°C (medium/high stringency), even more preferably at least 70°C (high stringency), and even more preferably at least 75°C (very high stringency).

Molecules to which the oligonucleotide probe hybridizes
5 under these conditions are detected using a x-ray film.

As previously noted, the isolated polynucleotides of the present invention include DNA and RNA. Methods for isolating DNA and RNA are well known in the art. DNA and RNA encoding genes of interest can be cloned in Gene Banks or DNA libraries
10 by means of methods known in the art.

Polynucleotides encoding polypeptides having endoglucanase activity of the invention are then identified and isolated by, for example, hybridization or PCR.

The present invention further provides counterpart
15 polypeptides and polynucleotides from different bacterial strains (orthologs or paralogs). Of particular interest are endoglucanase polypeptides from gram-positive alkalophilic strains, including species of *Bacillus*.

Species homologues of a polypeptide with endoglucanase
20 activity of the invention can be cloned using information and compositions provided by the present invention in combination with conventional cloning techniques. For example, a DNA sequence of the present invention can be cloned using chromosomal DNA obtained from a cell type that expresses the
25 protein. Suitable sources of DNA can be identified by probing Northern blots with probes designed from the sequences disclosed herein. A library is then prepared from chromosomal DNA of a positive cell line. A DNA sequence of the invention encoding an polypeptide having endoglucanase activity can then be isolated
30 by a variety of methods, such as by probing with probes designed from the sequences disclosed in the present specification and claims or with one or more sets of degenerate probes based on the disclosed sequences. A DNA sequence of the invention can also be cloned using the polymerase chain reaction, or PCR
35 (Mullis, U.S. Patent 4,683,202), using primers designed from the sequences disclosed herein. Within an additional method, the DNA library can be used to transform or transfect host cells, and expression of the DNA of interest can be detected with an

antibody (monoclonal or polyclonal) raised against the endoglucanase cloned from *B.licheniformis*, ATCC 14580, expressed and purified as described in Materials and Methods and Examples 1 and 3, or by an activity test relating to a polypeptide having
5 endoglucanase activity.

The endoglucanase encoding part of the DNA sequence cloned into plasmid pSJ1678 present in *Escherichia coli* DSM 12805 and/or an analogue DNA sequence of the invention may be cloned from a strain of the bacterial species *Bacillus licheniformis*,
10 preferably the strain ATCC 14580, producing the enzyme with endoglucanase activity, or another or related organism as described herein.

Alternatively, the analogous sequence may be constructed on the basis of the DNA sequence obtainable from the plasmid
15 present in *Escherichia coli* DSM 12805 (which is believed to be identical to the attached SEQ ID NO:1), e.g. be a sub-sequence thereof, and/or by introduction of nucleotide substitutions which do not give rise to another amino acid sequence of the endoglucanase encoded by the DNA sequence, but which corresponds
20 to the codon usage of the host organism intended for production of the enzyme, or by introduction of nucleotide substitutions which may give rise to a different amino acid sequence (i.e. a variant of the enzyme of the invention).

Alternatively, the DNA encoding an endoglucanase of the
25 invention may, in accordance with well-known procedures, conveniently be cloned from a suitable source, such as any of the below mentioned organisms, by use of synthetic oligonucleotide probes prepared on the basis of the DNA sequence obtainable from the plasmid present in *Escherichia coli* DSM
30 12805.

How to use a sequence of the invention to get other related sequences: The disclosed sequence information herein relating to a polynucleotide sequence encoding an endo-beta-1,4-glucanase of the invention can be used as a tool to identify
35 other homologous endoglucanases. For instance, polymerase chain reaction (PCR) can be used to amplify sequences encoding other

homologous mannanases from a variety of microbial sources, in particular of different *Bacillus* species.

POLYPEPTIDES:

5 The sequence of amino acids in position 26 to about position 485 of SEQ ID NO: 2 is a mature endoglucanase sequence of the catalytic active domain. The enzyme further comprises a cellulose binding domain (CBD) which is operably linked to the catalytic active domain and which is represented by an amino
10 acid sequence corresponding to from about position 485 to position 646 of SEQ ID NO:2. The CBD of the present endoglucanase belongs to family 3b, cf. below.

 The present invention also provides endoglucanase polypeptides that are substantially homologous to the
15 polypeptide of SEQ ID NO:2 and species homologs (paralogs or orthologs) thereof. The term "substantially homologous" is used herein to denote polypeptides having 75%, preferably at least 80%, more preferably at least 85%, and even more preferably at least 90%, sequence identity to the sequence shown in amino
20 acids nos. 26-485 or nos. 26-646 of SEQ ID NO:2 or their orthologs or paralogs. Such polypeptides will more preferably be at least 95% identical, and most preferably 98% or more identical to the sequence shown in amino acids nos. 26-646 of SEQ ID NO:2 or its orthologs or paralogs. Percent sequence
25 identity is determined by conventional methods, by means of computer programs known in the art such as GAP provided in the GCG program package (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711) as disclosed in Needleman,
30 S.B. and Wunsch, C.D., (1970), Journal of Molecular Biology, 48, 443-453, which is hereby incorporated by reference in its entirety. GAP is used with the following settings for polypeptide sequence comparison: GAP creation penalty of 3.0 and GAP extension penalty of 0.1.

35 Sequence identity of polynucleotide molecules is determined by similar methods using GAP with the following settings for DNA sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3.

Substantially homologous proteins and polypeptides are characterized as having one or more amino acid substitutions, deletions or additions. These changes are preferably of a minor nature, that is conservative amino acid substitutions (see Table 5 2) and other substitutions that do not significantly affect the folding or activity of the protein or polypeptide; small deletions, typically of one to about 30 amino acids; and small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, a small linker peptide of up to 10 about 20-25 residues, or a small extension that facilitates purification (an affinity tag), such as a poly-histidine tract, protein A (Nilsson et al., EMBO J. 4:1075, 1985; Nilsson et al., Methods Enzymol. 198:3, 1991. See, in general Ford et al., Protein Expression and Purification 2: 95-107, 1991, which is incorporated herein by reference. DNAs encoding affinity tags are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, NJ; New England Biolabs, Beverly, MA).

However, even though the changes described above preferably are of a minor nature, such changes may also be of a larger nature such as fusion of larger polypeptides of up to 300 amino acids or more both as amino- or carboxyl-terminal extensions to a polypeptide of the invention having endoglucanase activity.

Table 1

25 Conservative amino acid substitutions

	Basic:	arginine
		lysine
		histidine
30	Acidic:	glutamic acid
		aspartic acid
	Polar:	glutamine
		asparagine
35	Hydrophobic:	leucine
		isoleucine
		valine
	Aromatic:	phenylalanine
		tryptophan
		tyrosine

Small: glycine
 alanine
 serine
 threonine
5 methionine

In addition to the 20 standard amino acids, non-standard amino acids (such as 4-hydroxyproline, 6-*N*-methyl lysine, 2-aminoisobutyric acid, isovaline and α-methyl serine) may be substituted for amino acid residues of a polypeptide according to the invention. A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, and unnatural amino acids may be substituted for amino acid residues. "Unnatural amino acids" have been modified after protein synthesis, and/or have a chemical structure in their side chain(s) different from that of the standard amino acids. Unnatural amino acids can be chemically synthesized, or preferably, are commercially available, and include pipecolic acid, thiazolidine carboxylic acid, dehydroproline, 3- and 4-methylproline, and 3,3-dimethylproline.

Essential amino acids in the endoglucanase polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, Science 244: 1081-1085, 1989). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for biological activity (i.e. endoglucanase activity) to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton et al., J. Biol. Chem. 271:4699-4708, 1996. The active site of the enzyme or other biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos et al., Science 255:306-312, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992. The identities of essential amino acids can also be inferred

from analysis of homologies with polypeptides which are related to a polypeptide according to the invention.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis, recombination and/or shuffling followed by a relevant screening procedure, such as those disclosed by Reidhaar-Olson and Sauer (Science 241:53-57, 1988), Bowie and Sauer (Proc. Natl. Acad. Sci. USA 86:2152-2156, 1989), WO95/17413, or WO 95/22625. Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, or recombination/shuffling of different mutations (WO95/17413, WO95/22625), followed by selecting for functional a polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., Biochem. 30:10832-10837, 1991; Ladner et al., U.S. Patent No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., Gene 46:145, 1986; Ner et al., DNA 7:127, 1988).

Mutagenesis/shuffling methods as disclosed above can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides in host cells. Mutagenized DNA molecules that encode active polypeptides can be recovered from the host cells and rapidly sequenced using modern equipment. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide of interest, and can be applied to polypeptides of unknown structure.

Using the methods discussed above, one of ordinary skill in the art can identify and/or prepare a variety of polypeptides that are substantially homologous to residues 26 to about 485 or residues 26 to 646 of SEQ ID NO: 2 and retain the endoglucanase activity of the wild-type protein.

The endoglucanase enzyme of the invention may, in addition to the enzyme core comprising the catalytically domain, also comprise a cellulose binding domain (CBD), the cellulose binding domain and enzyme core (the catalytically active domain) of the enzyme being operably linked. The cellulose binding domain (CBD) may exist as an integral part the encoded enzyme as described

above and in the appended SEQ ID NO:2, or a CBD from another origin may be introduced into the endoglucanase thus creating an enzyme hybriide. In this context, the term "cellulose-binding domain" is intended to be understood as defined by Peter Tomme et al. "Cellulose-Binding Domains: Classification and Properties" in "Enzymatic Degradation of Insoluble Carbohydrates", John N. Saddler and Michael H. Penner (Eds.), ACS Symposium Series, No. 618, 1996. This definition classifies more than 120 cellulose-binding domains into 10 families (I-X), and demonstrates that CBDs are found in various enzymes such as cellulases (endoglucanases), xylanases, mannanases, arabinofuranosidases, acetyl esterases and chitinases. CBDs have also been found in algae, e.g. the red alga *Porphyra purpurea* as a non-hydrolytic polysaccharide-binding protein, see Tomme et al., *op.cit.* However, most of the CBDs are from cellulases and xylanases, CBDs are found at the N and C termini of proteins or are internal. Enzyme hybrids are known in the art, see e.g. WO 90/00609 and WO 95/16782, and may be prepared by transforming into a host cell a DNA construct comprising at least a fragment of DNA encoding the cellulose-binding domain ligated, with or without a linker, to a DNA sequence encoding the endoglucanase and growing the host cell to express the fused gene. Enzyme hybrids may be described by the following formula:

CBD - MR - X

wherein CBD is the N-terminal or the C-terminal region of an amino acid sequence corresponding to at least the cellulose-binding domain; MR is the middle region (the linker), and may be a bond, or a short linking group preferably of from about 2 to about 100 carbon atoms, more preferably of from 2 to 40 carbon atoms; or is preferably from about 2 to to about 100 amino acids, more preferably of from 2 to 40 amino acids; and X is an N-terminal or C-terminal region of a polypeptide encoded by the first or second DNA sequence of the invention.

In a preferred embodiment, the isolated polynucleotide molecule of the invention comprises a partial DNA sequence encoding a cellulose binding domain (CBD). An example of such a partial DNA sequence is the sequence corresponding to the

nucleotides in positions from about 485 to 1941 of SEQ ID NO:1 or the CBD encoding part of the DNA sequence cloned into the plasmid pSJ1678 present in *Escherichia coli*, DSM 12805. The isolated polynucleotide molecule of the invention may comprise a further partial nucleotide sequence encoding a linking region, the linking region operably linking the cellulose binding domain (CBD) and the catalytically active domain (CAD) of the enzyme encoded by the nucleotide sequence comprised by the isolated polynucleotide molecule. Preferably, the linking region consists of from about 2 amino acid residues to about 120 amino acid residues, especially 10-80 amino acid residues.

Immunological cross-reactivity

Polyclonal antibodies, especially monospecific polyclonal antibodies, to be used in determining immunological cross-reactivity may be prepared by use of a purified cellulolytic enzyme. More specifically, antiserum against the endoglucanase of the invention may be raised by immunizing rabbits (or other rodents) according to the procedure described by N. Axelsen et al. in: A Manual of Quantitative Immuno-electrophoresis, Blackwell Scientific Publications, 1973, Chapter 23, or A. Johnstone and R. Thorpe, Immunochemistry in Practice, Blackwell Scientific Publications, 1982 (more specifically p. 27-31). Purified immunoglobulins may be obtained from the antisera, for example by salt precipitation ($(\text{NH}_4)_2\text{SO}_4$), followed by dialysis and ion exchange chromatography, e.g. on DEAE-Sephadex.

Immunochemical characterization of proteins may be done either by Ouchterlony double-diffusion analysis (O. Ouchterlony in: Handbook of Experimental Immunology (D.M. Weir, Ed.), Blackwell Scientific Publications, 1967, pp. 655-706), by crossed immuno-electrophoresis (N. Axelsen et al., *supra*, Chapters 3 and 4), or by rocket immuno-electrophoresis (N. Axelsen et al., Chapter 2).

Microbial Sources

For the purpose of the present invention the term "obtained from" or "obtainable from" as used herein in

connection with a specific source, means that the enzyme is produced or can be produced by the specific source, or by a cell in which a gene from the source have been inserted.

It is at present contemplated that the cellulase of the invention may be obtained from a gram-positive bacterium belonging to a strain of the genus *Bacillus*, in particular a strain of *Bacillus licheniformis*.

In a preferred embodiment, the cellulase of the invention is obtained from the strain *Bacillus licheniformis*, ATCC 14580. It is at present contemplated that a DNA sequence encoding an enzyme homologous to the enzyme of the invention may be obtained from other strains belonging to the genus *Bacillus*.

An isolate of a strain of *Bacillus licheniformis* from which an endo- β -1,4-glucanase of the invention can be derived is publicly available from American Type Culture Collection (ATCC) under the deposition number ATCC 14580.

Further, the plasmid pSJ1678 comprising the DNA sequence encoding the endoglucanase of the invention has been transformed into a strain of the *Escherichia coli* and deposited under the deposition number DSM 12805.

Recombinant expression vectors

A recombinant vector comprising a DNA construct encoding the enzyme of the invention may be any vector, which may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector, which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome in part or in its entirety and replicated together with the chromosome(s) into which it has been integrated.

The vector is preferably an expression vector in which the DNA sequence encoding the enzyme of the invention is operably

linked to additional segments required for transcription of the DNA. In general, the expression vector is derived from plasmid or viral DNA, or may contain elements of both. The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through the DNA sequence coding for the enzyme.

The promoter may be any DNA sequence, which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell.

Examples of suitable promoters for use in bacterial host cells include the promoter of the *Bacillus stearothermophilus* maltogenic amylase gene, the *Bacillus licheniformis* alpha-amylase gene, the *Bacillus amyloliquefaciens* alpha-amylase gene, the *Bacillus subtilis* alkaline protease gen, or the *Bacillus pumilus* xylosidase gene, or the phage Lambda P_H or P_L promoters or the *E. coli* lac, trp or tac promoters.

The DNA sequence encoding the enzyme of the invention may also, if necessary, be operably connected to a suitable terminator.

The recombinant vector of the invention may further comprise a DNA sequence enabling the vector to replicate in the host cell in question.

The vector may also comprise a selectable marker, e.g. a gene the product of which complements a defect in the host cell, or a gene encoding resistance to e.g. antibiotics like kanamycin, chloramphenicol, erythromycin, tetracycline, spectinomycine, or the like, or resistance to heavy metals or herbicides.

To direct an enzyme of the present invention into the secretory pathway of the host cells, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) may be provided in the recombinant vector. The secretory signal sequence is joined to the DNA sequence encoding the enzyme in the correct reading frame. Secretory signal sequences are commonly positioned 5' to the DNA sequence

encoding the enzyme. The secretory signal sequence may be that normally associated with the enzyme or may be from a gene encoding another secreted protein.

The procedures used to ligate the DNA sequences coding for the present enzyme, the promoter and optionally the terminator and/or secretory signal sequence, respectively, or to assemble these sequences by suitable PCR amplification schemes, and to insert them into suitable vectors containing the information necessary for replication or integration, are well known to persons skilled in the art (cf., for instance, Sambrook et al., op.cit.).

Host cells

The cloned DNA molecule introduced into the host cell may be either homologous or heterologous to the host in question. If homologous to the host cell, i.e. produced by the host cell in nature, it will typically be operably connected to another promoter sequence or, if applicable, another secretory signal sequence and/or terminator sequence than in its natural environment. The term "homologous" is intended to include a DNA sequence encoding an enzyme native to the host organism in question. The term "heterologous" is intended to include a DNA sequence not expressed by the host cell in nature. Thus, the DNA sequence may be from another organism, or it may be a synthetic sequence.

The host cell into which the cloned DNA molecule or the recombinant vector of the invention is introduced may be any cell which is capable of producing the desired enzyme and includes bacteria, yeast, fungi and higher eukaryotic cells.

Examples of bacterial host cells which on cultivation are capable of producing the enzyme of the invention may be a gram-positive bacteria such as a strain of *Bacillus*, in particular *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus megatherium*, *Bacillus stearothermophilus*, *Bacillus subtilis* and *Bacillus thuringiensis*, a strain of *Lactobacillus*, a strain of

Streptococcus, a strain of *Streptomyces*, in particular *Streptomyces lividans* and *Streptomyces murinus*, or the host cell may be a gram-negative bacteria such as a strain of *Escherichia coli*.

5 The transformation of the bacteria may be effected by protoplast transformation, electroporation, conjugation, or by using competent cells in a manner known per se (cf. e.g. Sambrook et al., supra).

10 When expressing the enzyme in a bacterium such as *Escherichia coli*, the enzyme may be retained in the cytoplasm, typically as insoluble granules (known as inclusion bodies), or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed and the granules are recovered and denatured after which the enzyme
15 is refolded by diluting the denaturing agent. In the latter case, the enzyme may be recovered from the periplasmic space by disrupting the cells, e.g. by sonication or osmotic shock, to release the contents of the periplasmic space and recovering the enzyme.

20 When expressing the enzyme in a gram-positive bacterium such as a strain of *Bacillus* or a strain of *Streptomyces*, the enzyme may be retained in the cytoplasm, or may be directed to the extracellular medium by a bacterial secretion sequence.

25 Examples of a fungal host cell which on cultivation are capable of producing the enzyme of the invention is e.g. a strain of *Aspergillus* or *Fusarium*, in particular *Aspergillus awamori*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, and *Fusarium oxysporum*, and a strain of *Trichoderma*, preferably *Trichoderma harzianum*, *Trichoderma reesei* and
30 *Trichoderma viride*.

 Fungal cells may be transformed by a process involving protoplast formation and transformation of the protoplasts followed by regeneration of the cell wall in a manner known per se. The use of a strain of *Aspergillus* as a host cell is

described in EP 238 023 (Novo Nordisk A/S), the contents of which are hereby incorporated by reference.

Examples of a host cell of yeast origin which on cultivation are capable of producing the enzyme of the invention
5 is e.g. a strain of *Hansenula* sp., a strain of *Kluyveromyces* sp., in particular *Kluyveromyces lactis* and *Kluyveromyces marcianus*, a strain of *Pichia* sp., a strain of *Saccharomyces*, in particular *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces kluyveri* and *Saccharomyces uvarum*, a
10 strain of *Schizosaccharomyces* sp., in particular *Schizosaccharomyces pombe*, and a strain of *Yarrowia* sp., in particular *Yarrowia lipolytica*.

Examples of a host cell of plant origin which on cultivation are capable of producing the enzyme of the invention
15 is e.g. a plant cell of *Solanum tuberosum* or *Nicotiana tabacum*.

Method of producing a cellulolytic enzyme

The present invention provides a method of producing an isolated enzyme according to the invention, wherein a suitable
20 host cell, which has been transformed with a DNA sequence encoding the enzyme, is cultured under conditions permitting the production of the enzyme, and the resulting enzyme is recovered from the culture.

As defined herein, an isolated polypeptide (e.g. an
25 enzyme) is a polypeptide which is essentially free of other polypeptides, e.g., at least about 20% pure, preferably at least about 40% pure, more preferably about 60% pure, even more preferably about 80% pure, most preferably about 90% pure, and even most preferably about 95% pure, as determined by SDS-PAGE.

30 The term "isolated polypeptide" may alternatively be termed "purified polypeptide".

When an expression vector comprising a DNA sequence encoding the enzyme is transformed into a heterologous host cell it is possible to enable heterologous recombinant production of
35 the enzyme of the invention.

Thereby it is possible to make a highly purified or monocomponent cellulolytic composition, characterized in being free from homologous impurities.

In this context, homologous impurities mean any impurities (e.g. other polypeptides than the enzyme of the invention) which originate from the homologous cell where the enzyme of the invention is originally obtained from.

In the present invention the homologous host cell may be a strain of *Bacillus licheniformis*.

The medium used to culture the transformed host cells may be any conventional medium suitable for growing the host cells in question. The expressed cellulolytic enzyme may conveniently be secreted into the culture medium and may be recovered therefrom by well-known procedures including separating the cells from the medium by centrifugation or filtration, precipitating proteinaceous components of the medium by means of a salt such as ammonium sulphate, followed by chromatographic procedures such as ion exchange chromatography, affinity chromatography, or the like.

20

Enzyme compositions

In a still further aspect, the present invention relates to an enzyme composition comprising an enzyme exhibiting endoglucanase activity as described above.

The enzyme composition of the invention may, in addition to the endoglucanase of the invention, comprise one or more other enzyme types, for instance hemicellulase such as xylanase and mannanase, other cellulase or endo- β -1,4-glucanase components, chitinase, lipase, esterase, pectinase, cutinase, phytase, oxidoreductase (peroxidase, haloperoxidase, oxidase, laccase), protease, amylase, reductase, phenoloxidase, ligninase, pullulanase, pectate lyase, xyloglucanase, pectin acetyl esterase, polygalacturonase, rhamnogalacturonase, pectin lyase, pectin methylesterase, cellobiohydrolase, transglutaminase; or mixtures thereof.

35

The enzyme composition may be prepared in accordance with methods known in the art and may be in the form of a liquid or a

dry composition. For instance, the enzyme composition may be in the form of a granulate or a micro-granulate. The enzyme to be included in the composition may be stabilized in accordance with methods known in the art.

5 Endoglucanases have potential uses in a lot of different industries and applications. Examples are given below of preferred uses of the enzyme composition of the invention. The dosage of the enzyme composition of the invention and other conditions under which the composition is used may be determined on
10 the basis of methods known in the art.

The enzyme composition according to the invention may be useful for at least one of the following purposes.

The enzyme

15 In a preferred embodiment of the present invention, the endoglucanase exhibits activity at a pH in the range of 4-11, preferably 5.5-10.5.

Uses

20 Biomass degradation

The enzyme or the enzyme composition according to the invention may be applied advantageously e.g. as follows:

- For debarking, i.e. pretreatment with hydrolytic enzymes which may partly degrade the pectin-rich cambium layer
25 prior to debarking in mechanical drums resulting in advantageous energy savings.

- For defibration (refining or beating), i.e. treatment of material containing cellulosic fibers with hydrolytic enzymes prior to the refining or beating which results in reduction of
30 the energy consumption due to the hydrolysing effect of the enzymes on the surfaces of the fibers.

- For fibre modification, i.e. improvement of fibre properties where partial hydrolysis across the fibre wall is needed which requires deeper penetrating enzymes (e.g. in order
35 to make coarse fibers more flexible).

- For drainage: The drainability of papermaking pulps may be improved by treatment of the pulp with hydrolysing

enzymes. Use of the enzyme or enzyme composition of to the invention may be more effective, e.g. result in a higher degree of loosening bundles of strongly hydrated micro-fibrils in the fines fraction that limits the rate of drainage by blocking
5 hollow spaces between the fibers and in the wire mesh of the paper machine.

The treatment of lignocellulosic pulp may, e.g., be performed as described in WO 93/08275, WO 91/02839 and WO 92/03608.

10

Use in the detergent industry

The enzyme or enzyme composition of the invention may be useful in a detergent composition for house-hold or industrial laundering of textiles and garment, and to a process for machine
15 treatment of fabrics comprising treating fabric during a washing cycle of a machine washing process with a washing solution containing the enzyme or enzyme preparation of the invention.

Typically, the detergent composition of the invention comprises conventional ingredients such as surfactants (anionic,
20 nonionic, zwitterionic, amphoteric), builders, and other ingredients, e.g. as described in WO 97/01629 which is hereby incorporated by reference.

Textile applications

25 In another embodiment, the present invention relates to use of the endoglucanase of the invention in the bio-polishing process. Bio-Polishing is a specific treatment of the yarn surface which improves fabric quality with respect to handle and appearance without loss of fabric wettability. The most impor-
30 tant effects of Bio-Polishing can be characterized by less fuzz and pilling, increased gloss/luster, improved fabric handle, increased durable softness and altered water absorbency. Bio-Polishing usually takes place in the wet processing of the manufacture of knitted and woven fabrics. Wet processing
35 comprises such steps as e.g. desizing, scouring, bleaching, washing, dying/printing and finishing. During each of these steps, the fabric is more or less subjected to mechanical

action. In general, after the textiles have been knitted or woven, the fabric proceeds to a desizing stage, followed by a scouring stage, etc. Desizing is the act of removing size from textiles. Prior to weaving on mechanical looms, warp yarns are
5 often coated with size starch or starch derivatives in order to increase their tensile strength. After weaving, the size coating must be removed before further processing the fabric in order to ensure a homogeneous and wash-proof result. It is known that in order to achieve the effects of Bio-Polishing, a
10 combination of cellulytic and mechanical action is required. It is also known that "super-softness" is achievable when the treatment with a cellulase is combined with a conventional treatment with softening agents. It is contemplated that use of the endoglucanase of the invention for bio-polishing of cellulo-
15 sic fabrics is advantageous, e.g. a more thorough polishing can be achieved. Bio-polishing may be obtained by applying the method described e.g. in WO 93/20278.

Stone-washing

20 It is known to provide a "stone-washed" look (localized abrasion of the colour) in dyed fabric, especially in denim fabric or jeans, either by washing the denim or jeans made from such fabric in the presence of pumice stones to provide the desired localized lightening of the colour of the fabric or by
25 treating the fabric enzymatically, in particular with cellulytic enzymes. The treatment with an endoglucanase of the present invention may be carried out either alone such as disclosed in US 4,832,864, together with a smaller amount of pumice than required in the traditional process, or together with perlite
30 such as disclosed in WO 95/09225.

DETERMINATION OF CMC UNITS

CMC units is determined using 0.1 M Mops buffer pH 7.5 at 60°C. 20 min incubation and determination of the formation of
35 reducing sugars using PHAB. One CMC unit corresponds to the formation of 1 micromole glucose equivalent per min. The CMC (Car-

boxy Methyl Cellulose 7L from Hercules) final concentration is 0.75%, DS 0.7.

MATERIALS AND METHODS

5 Strains

Bacillus licheniformis ATCC 14580.

B. subtilis PL2306. This strain is the *B. subtilis* DN1885 with disrupted *apr* and *npr* genes (Diderichsen et al. (1990)) disrupted in the transcriptional unit of the known *Bacillus subtilis* cellulase gene, resulting in cellulase negative cells. The disruption was performed essentially as described in A.L. Sonenshein et al. (1993).

Competent cells were prepared and transformed as described by Yasbin et al. (1975).

15

Plasmids

pSJ1678 disclosed in International Patent Publication WO 94/19454.

20 pMOL944:

This plasmid is a pUB110 derivative essentially containing elements making the plasmid propagatable in *Bacillus subtilis*, kanamycin resistance gene and having a strong promoter and signal peptide cloned from the *amyL* gene of *B. licheniformis* ATCC14580. The signal peptide contains a *SacII* site making it convenient to clone the DNA encoding the mature part of a protein in-fusion with the signal peptide. This results in the expression of a Pre-protein, which is directed towards the exterior of the cell.

30 The plasmid was constructed by means of conventional genetic engineering techniques, which are briefly described in the following.

Construction of pMOL944:

The pUB110 plasmid (McKenzie, T. et al., 1986) was digested with the unique restriction enzyme NciI. A PCR fragment amplified from the amyL promoter encoded on the plasmid pDN1981 (Jørgensen P.L. et al. (1990)) was digested with NciI and inserted in the NciI digested pUB110 to give the plasmid pSJ2624.

The two PCR primers used have the following sequences:

LWN5494 5'-GTCGCCGGGGCGGCCGCTATCAATTGGTAACTGTATCTCAGC -3'
LWN5495 5'-GTCGCCGGGGCAGCTCTGATCAGGTACCAAGCTTGTGACCTGCAGAA
TGAGGCAGCAAGAAGAT -3'

The primer #LWN5494 inserts a NotI site in the plasmid.

The plasmid pSJ2624 was then digested with SacI and NotI and a new PCR fragment amplified on amyL promoter encoded on the pDN1981 was digested with SacI and NotI and this DNA fragment was inserted in the SacI-NotI digested pSJ2624 to give the plasmid pSJ2670.

This cloning replaces the first amyL promoter cloning with the same promoter but in the opposite direction. The two primers used for PCR amplification have the following sequences:

#LWN5938 5'-GTCGCCGGGGCGCTGATCACGTACCAAGCTTGTGACCTGCAGAATG
AGGCAGCAAGAAGAT -3'
#LWN5939 5'-GTCGGAGCTCTATCAATTGGTAACTGTATCTCAGC -3'

The plasmid pSJ2670 was digested with the restriction enzymes PstI and BclI and a PCR fragment amplified from a cloned DNA sequence encoding the alkaline amylase SP722 (disclosed in the International Patent Application published as WO95/26397 which is hereby incorporated by reference in its entirety) was digested with PstI and BclI and inserted to give the plasmid pMOL944. The two primers used for PCR amplification have the following sequence:

#LWN7864 5'-AACAGCTGATCAGACTGATCTTTTAGCTTGGCAC-3'
#LWN7901 5'-AACTGCAGCCGCGGCACATCATAATGGGACAAATGGG -3'

The primer #LWN7901 inserts a SacII site in the plasmid.

General molecular biology methods

Unless otherwise mentioned the DNA manipulations and transformations were performed using standard methods of

molecular biology (Sambrook et al. (1989); Ausubel, F. M. et al. (eds) (1995); Harwood, C. R., and Cutting, S. M. (eds.) (1990)).

Enzymes for DNA manipulations were used according to the specifications of the suppliers (e.g. restriction endonucleases, ligases etc. are obtainable from New England Biolabs, Inc.).

Media

TY (as described in Ausubel et al. (1995)).

LB agar (as described in Ausubel et al. (1995)).

10 **LBPG** is LB agar supplemented with 0.5% glucose and 0.05 M potassium phosphate, pH 7.0

BPX media is described in EP 0 506 780 (WO 91/09129).

The following examples illustrate the invention.

15

EXAMPLE 1

Cloning and expression of endo-beta-1,4-glucanase from *Bacillus licheniformis*

Genomic DNA preparation

20 Strain *Bacillus licheniformis* ATCC 14580 was propagated in liquid medium 3 as specified by ATCC (American Type Culture Collection, USA). After 18 hours incubation at 37°C and 300 rpm, the cells were harvested, and genomic DNA isolated by the method described by Pitcher et al. (1989).

25

Genomic Library Construction

Genomic DNA of *Bacillus licheniformis* ATCC 14580 was partially digested with restriction enzyme Sau3A and size-fractionated by electrophoresis on a 0.7 % agarose gel. Fragments of between 2 and 7 kb in size were isolated by electrophoresis onto DEAE-cellulose paper (Dretzen et al. (1981)). Isolated DNA fragments were ligated to BamHI digested pSJ1678 plasmid DNA.

Ligated DNA was used in electroporation of *E. coli* SJ2, the transformed cells were plated on LB-agar plates containing 10 mg/ml Chloramphenicol and 0.1% CMC (Sodium-Carboxy-Methyl-Cellulose, Aqualon, France), the plates were incubated 18 hours
5 at 37 °C.

Identification of positive clones by colony hybridization

A DNA library in *E. coli*, constructed as described above, was screened on LB agar plates containing 0.1% CMC (Sodium-Carboxy-Methyl-Cellulose, Aqualon, France) and 10 µg/ml
10 Chloramphenicol and incubated overnight at 37°C. The transformants were subsequently replica plated onto the same type of plates, and these new plates were incubated 8 hours or overnight at 37°C.

15 The original plates were coloured using 25 ml of a aqueous solution containing 1 mg/ml of Congo Red (SIGMA, USA). The colouring was continued for half an hour with moderate orbital shaking, after which the plates were washed two times 15 minutes using 1 M NaCl.

20 Yellowish halos appeared at positions where cellulase positive clones were present, from the replica plates these cellulase positive clones were rescued and re-streaked onto LB agar plates containing 0.1% CMC and 9 µg/ml Chloramphenicol and incubated overnight at 37°C.

25

Characterization of positive clones

From the re-streaking plates the endoglucanase positive clones were obtained as single colonies, and plasmids were extracted. Phenotypes were confirmed by retransformation of *E. coli*
30 SJ2, and plasmids characterized by restriction digests. One positive clone was termed MB629-3.

The endoglucanase gene was characterized by DNA sequencing using the Taq deoxy-terminal cycle sequencing kit (Perkin-Elmer, USA) and performing primer walking, starting with primers unique
35 to the pSJ1678 plasmid and on each side of the cloned endoglucanase encoding DNA fragment.

Analysis of the sequence data was performed according to Devereux *et al.* (1984). The sequence corresponds to the DNA sequence shown in SEQ ID NO: 1.

The DNA sequence of the invention coding for the family 9 endo-beta-1,4-glucanase represented by amino acid sequence SEQ ID NO:2 (also denoted Cel9) was PCR amplified using the PCR primer set consisting of these two oligo nucleotides:

Cel9.B.lich.upper.PstI

10 5'-CAT CAT TCT GCA GCC GCG GCA GCT TCT GCT GAA GAA TAT CCT C-3' 7

Cel9.B.lich.lower.NotI

5'-GCG AGA ATA GCG GCC GCT AGT AAC CGG GCT CAT GTC CG-3'

15 Restriction sites PstI and NotI are underlined.

Chromosomal DNA isolated from *B.licheniformis* ATCC 14580 as described above was used as template in a PCR reaction using Amplitaq DNA Polymerase (Perkin Elmer) according to manufacturers instructions. The PCR reaction was set up in PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.01 % (w/v) gelatin) containing 200 μM of each dNTP, 2.5 units of AmpliTag polymerase (Perkin-Elmer, Cetus, USA) and 100 pmol of each primer.

The PCR reactions were performed using a DNA thermal cycler (Landgraf, Germany). One incubation at 94°C for 1 min followed by thirty cycles of PCR performed using a cycle profile of denaturation at 94°C for 30 sec, annealing at 60°C for 1 min, and extension at 72°C for 2 min. Five-μl aliquots of the amplification product was analyzed by electrophoresis in 0.7 % agarose gels (NuSieve, FMC). The appearance of a DNA fragment size 2.0 kb indicated proper amplification of the gene segment.

Subcloning of PCR fragment

Fortyfive-μl aliquots of the PCR products generated as described above were purified using QIAquick PCR purification kit (Qiagen, USA) according to the manufacturer's instructions. The purified DNA was eluted in 50 μl of 10mM Tris-HCl, pH 8.5.

5 μ g of pMOL944 and twentyfive- μ l of the purified PCR fragment was digested with PstI and NotI, electrophoresed in 0.8 % low gelling temperature agarose (SeaPlaque GTG, FMC) gels, the relevant fragments were excised from the gels, and purified using
5 QIAquick Gel extraction Kit (Qiagen, USA) according to the manufacturer's instructions. The isolated PCR DNA fragment was then ligated to the PstI-NotI digested and purified pMOL944. The ligation was performed overnight at 16°C using 0.5 μ g of each DNA fragment, 1 U of T4 DNA ligase and T4 ligase buffer (Boe-
10 hringer Mannheim, Germany).

The ligation mixture was used to transform competent *B.subtilis* PL2306. The transformed cells were plated onto LBPG- 10 μ g/ml of Kanamycin plates. After 18 hours incubation at 37°C several clones were re-streaked on fresh agar plates and also
15 grown in liquid TY cultures with 10 μ g/ ml kanamycin and incubated overnight at 37°C. Next day 1 ml of cells were used to isolate plasmid from the cells using the Qiaprep Spin Plasmid Miniprep Kit #27106 according to the manufacturers recommendations for *B.subtilis* plasmid preparations. This plasmid DNA was
20 used as template for DNA sequencing.

One clone containing the endo-beta-1,4-glucanase gene of the invention was kept, this clone was denoted MB905.

The plasmid from MB905 was introduced to a derivative of *B.licheniformis* ATCC 14580, for expression trials. This strain
25 was termed MB924. The cloned DNA sequence was expressed in *B.licheniformis* by fermenting the cells in BP-X media at 37°C for 5 days at 300 rpm. The endoglucanase protein that appeared in the supernatant corresponded to the mature protein of SEQ ID NO:2, ie comprised the protein sequence corresponding to the
30 amino acids at position 26-646 of SEQ ID NO:2.

EXAMPLE 2

Purification and characterization of endo-beta-1,4-glucanase from *Bacillus licheniformis*

Purification

MB924 obtained as described in example 1 was grown in 15 x 200 ml BPX media with 10 µg/ml of Kanamycin in 500 ml two baffled shake flasks for 5 days at 37°C at 300 rpm, whereby 2500 ml of culture broth was obtained. The culture fluid was diluted with one volume of ionized water and pH adjusted to 7.5, using acetic acid. Then 112.5 ml of cationic agent (C521 10%) and 225 ml of anionic agent (A130 0.1%) was added during agitation for flocculation. The flocculated material was separated by centrifugation using a Sorval RC 3B centrifuge at 10000 rpm for 30 min at 6°C. The resulting supernatant contained 120 CMCunits per ml in a total volume of 5000 ml.

The supernatant was clarified using Whatman glass filters GF/D and C and finally concentrated on a filtron UF membrane with a cut off of 10 kDa. The total volume of 1750 ml was adjusted to pH 8.0.

For obtaining a highly purified endoglucanase a final step using Q-sepharose anion-exchange chromatography was carried out. 1750 ml of the solution was applied to a 800 ml column containing Q-Sepharose (Pharmacia) equilibrated with a buffer of 50 mmol Tris pH 8.0. The endoglucanase bound and was eluted using a 0.5 M NaCl gradient. The more than 95% of the endoglucanase was concentrated.

Characterisation

The pure enzyme gave a single band in SDS-PAGE of 67 kDa and an isoelectric point of around 5.6.

The protein concentration was determined using a molar extinction coefficient of 171640 (based on the amino acid composition deducted from the sequence). The pH activity profiles showed more than 50% relative activity between pH 6.0 and 9.2. at 60°. The temperature optimum was 65° at pH 7.5. DSC showed melting at 77° C at pH 6.2.

N-terminal determination of the pure endoglucanase:

EYPHNYALLQK

The pure endoglucanase comprises a catalytic domain belonging to family 9 of glycosyl hydrolases, which domain corre-

sponds to the amino acid sequence from about position 26 to about position 485 of SEQ ID NO:2, and a cellulase binding domain (CBD) which is linked to the catalytic domain and is represented by the the amino acid sequence from about position 486 to
5 position 644 of SEQ ID NO:2. The CBD belongs to family 3b.

Immunological properties: At the Danish company DAKO, rabbit polyclonal monospecific serum was raised against the highly purified endo-beta-1,4-glucanase using conventional techniques. The serum formed a nice single precipitate in agarose gels with
10 the endoglucanase of the invention.

LITERATURE

- Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology". John Wiley and Sons, 1995.
- Axelsen, N., et al. in: A Manual of Quantitative Immuno-electrophoresis, Blackwell Scientific Publications, 1973, Chapter 23.
- Denman, S. et al.: Characterization of a *Neocallumastix patriarum* cellulase cDNA (celA) homologous to *Trichoderma reesei* cellobiohydrolase II, *Appl. Environ. Microbiol.* (1996), 62(6), 1889-1896.
- Damude, H.G. et al.: Substrate specificity of endoglucanase A from *Cellulomonas fimi*: fundamental differences between endoglucanases and exoglucanases from family 6, *Biochem.J.* (1996), 315(2), 467-72.
- Devereux et al. (1984) *Nucleic Acids Res.* 12, 387-395.
- Diderichsen, B., Wedsted, U., Hedegaard, L., Jensen, B. R., Sjøholm, C. (1990) Cloning of aldB, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. *J. Bacteriol.*, 172, 4315-4321
- Dretzen, G., Bellard, M., Sassone-Corsi, P., Chambon, P. (1981) A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Anal. Biochem.*, 112, 295-298.
- Feinberg, A. P. and Vogelstein, B. (1983) *Anal. Biochem.* 132:6-13.
- Gilbert, H.J. and Hazlewood, G.P. (1993) *J. Gen. Microbiol.* 139:187-194.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller Jr., R.C. and Warren, R.A.J.: Domains in microbial β -1,4-glycanases; sequence conservation, function, and enzyme families. Microbiol. Rev. 55 (1991), 305-315.

Harwood, C. R., and Cutting, S. M. (eds.) "Molecular Biological Methods for Bacillus". John Wiley and Sons, 1990.

Henrissat, B.: A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 280 (1991), 309-316.

Henrissat, B., and Bairoch, A.: New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 293 (1993), 781-788.

Johnstone, A. and R. Thorpe, Immunochemistry in Practice, Blackwell Scientific Publications, 1982 (pp. 27-31).

Jørgensen P.L. et al. (1990) Gene, 96, p.37-41.

Leatherbarrow, R. J. (1992) Grafit version 3.0 Erithacus Software Ltd. Staines, U.K.

Lever, M. (1972) A new reaction for colormetric determination of carbohydrates. Anal. Biochem. 47, 273-279.

McKenzie, T. et al., 1986, Plasmid 15:93-103.

O. Ouchterlony in: Handbook of Experimental Immunology (D.M. Weir, Ed.), Blackwell Scientific Publications, 1967, pp. 655-706.

Pitcher, D. G., Saunders, N. A., Owen, R. J. (1989): Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett. Appl. Microbiol., 8, 151-156.

Quillet, L. et al.: The gene encoding the beta-1,4-endoglucanase (Cel) from *Myxococcus xanthus*: evidence for independent acquisition by horizontal transfer of binding and catalytic domains from actinomycetes, *Gene* (1995), 158(1), 23-9.

Sambrook et al. (1989) *Molecular cloning: A laboratory manual*, Cold Spring Harbor lab., Cold Spring Harbor, NY.

- 5 Eds. A.L. Sonenshein, J.A. Hoch and Richard Losick (1993) *Bacillus subtilis* and other Gram-Positive Bacteria, American Society for microbiology, p.618.

- 10 Yasbin, R.E., Wilson, G.A. and Young, F.E. (1975) Transformation and transfection in lysogenic strains of *Bacillus subtilis*: evidence for selective induction of prophage in competent cells. *J. Bacteriol.* 121:296-304.